Giới thiệu Xác suất (Introduction to Probability)

Thống kê máy tính và ứng dụng (ĐTTX)

Vũ Quốc Hoàng (vqhoang@fit.hcmus.edu.vn)
FIT - HCMUS

Ngày 20 tháng 2 năm 2022

Nội dung

1. Thí nghiệm ngẫu nhiên
2. Biến cố
3. Xác suất
4. Mô hình xác suất đơn giản
5. Công thức cộng xác suất
6. Kỹ thuật đếm
7. Phương pháp xấp xỉ xác suất bằng mô phỏng
8. Ví dụ minh họa - Bài toán sinh nhật

Nội dung

1. Thí nghiệm ngẫu nhiên
2. Biến cố
3. Xác suất
4. Mô hình xác suất đơn giản
5. Công thức cộng xác suất
6. Kỹ thuật đếm
7. Phương pháp xấp xỉ xác suất bằng mô phỏng
8. Ví dụ minh họa - Bài toán sinh nhật

Thí nghiệm ngẫu nhiên

- **Thí nghiệm ngẫu nhiên** (random experiment) là các quá trình/hoạt động/thử nghiệm/công việc/thao tác
  - không biết chắc chắn **kết quả** (outcome),
  - nhưng xác định được tập tất cả các kết quả có thể.
- Tập tất cả các kết quả có thể của một thí nghiệm được gọi là **không gian mẫu** (sample space) của thí nghiệm, thường được kí hiệu là S hay Ω (omega).
- Bước đầu tiên của việc khảo sát một thí nghiệm là xác định không gian mẫu một cách "đầy đủ", "đúng đắn" và "tiện lợi".

Thí nghiệm ngẫu nhiên - Ví dụ

- Tung một đồng xu

$$\Omega = \{\text{Mặt ngửa}, \text{Mặt sấp}\} = \{N, S\} = \{H, T\} = \{0, 1\}.$$

- Tung một xúc xắc

$$\Omega = \{1, 2, 3, 4, 5, 6\}.$$

- Tung đồng thời 2 xúc xắc hoặc tung một xúc xắc 2 lần

$$\Omega = \{(i, j) : i, j \in \{1, 2, 3, 4, 5, 6\}\} = \{1, ..., 6\}^2.$$

- Tung một đồng xu 3 lần

$$\Omega_1 = \{0, 1, 2, 3\} \text{ (số lần được ngửa).}$$

$$\Omega_2 = \{HHH, HHT, ..., TTT\} = \{H, T\}^3 \text{ (chi tiết từng lần),}$$

Thí nghiệm ngẫu nhiên - Ví dụ (tt)

- Tung một đồng xu cho đến khi ra mặt sấp thì dừng

$$\Omega = \{T, HT, HHT, HHHT, ...\}.$$

- Đo nhiệt độ tại một địa điểm

$$\Omega_1 = \mathbb{R} = (-\infty, \infty) \text{ (độ C)},$$

$$\Omega_2 = [-273.15, \infty) \text{ (độ C)},$$

$$\Omega_3 = [-100, 3.6 \times 10^9] \text{ (độ C)},$$

$$\Omega_4 = [0, 1] \text{ (độ Planck)}.$$

- Đo chiều cao của một người

$$\Omega_1 = [0, \infty) \text{ (mét)},$$

$$\Omega_2 = [0.55, 2.51] \text{ (mét)}.$$

Thí nghiệm ngẫu nhiên

- Không gian mẫu **rời rạc** (discrete)
  - **Hữu hạn** (finite): có tương ứng 1-1 giữa Ω và tập $\{1, 2, ..., n\}, |\Omega| = n < \infty$,
  - **Vô hạn đếm được** (countable): có tương ứng 1-1 giữa Ω và $\mathbb{N} = \{1, 2, ...\}$.
- Không gian mẫu **liên tục** (continuous): khoảng con của $\mathbb{R}$.

Ví dụ

- Hữu hạn: Tung một đồng xu k lần

$$\Omega = \{H, T\}^k, |\Omega| = 2^k.$$

- Vô hạn đếm được: Tung một đồng xu cho đến khi ra mặt sấp thì dừng

$$\Omega = \{T, HT, HHT, HHHT, ...\}.$$

- Liên tục: Đo chiều cao của một người

$$\Omega_1 = [0, \infty) \text{ (mét)}.$$

- Nếu chiều cao được đo với độ chính xác đến cm thì sao?

Nội dung

1. Thí nghiệm ngẫu nhiên
2. Biến cố
3. Xác suất
4. Mô hình xác suất đơn giản
5. Công thức cộng xác suất
6. Kỹ thuật đếm
7. Phương pháp xấp xỉ xác suất bằng mô phỏng
8. Ví dụ minh họa - Bài toán sinh nhật

Biến cố

- Nếu việc xảy ra hay không của một tình huống E được xác định hoàn toàn khi biết kết quả ω của thí nghiệm T thì E được gọi là **biến cố** (event) **liên quan** đến T
  - Nếu ω làm cho E xảy ra thì ω được gọi là một **kết quả thuận lợi** cho E.
  - Một biến cố được đặc trưng bởi tập các kết quả thuận lợi cho nó.
- Biến cố là tập con của không gian mẫu

$$E = \{\omega \in \Omega : \omega \text{ thuận lợi cho } E\} \subset \Omega.$$

- Với mỗi kết quả $\omega \in \Omega$ ta đồng nhất $\{\omega\}$ với ω và gọi ω là **biến cố sơ cấp** (elementary event).

Biến cố - Ví dụ

Tung một đồng xu, $\Omega = \{H, T\}$, chỉ có 4 biến cố liên quan

- $\{\} = \varnothing$: được cả hai mặt, không ra mặt nào, ...
- $\{H\} \equiv H$: biến cố sơ cấp "được mặt ngửa".
- $\{T\} \equiv T$: biến cố sơ cấp "được mặt sấp".
- $\{H, T\} = \Omega$: được một trong hai mặt, được ít nhất một mặt, ...

Tung một đồng xu 3 lần

- $\Omega_1 = \{0, 1, 2, 3\}$
  - "Được 2 lần ngửa": $\{2\}$.
  - "Cả 3 lần đều ngửa": $\{3\}$.
  - "Lần 1 và lần 2 được ngửa": không là biến cố liên quan.
- $\Omega_2 = \{H, T\}^3$
  - "Được 2 lần ngửa": $\{HHT, HTH, THH\}$.
  - "Cả 3 lần đều ngửa": $\{HHH\}$.
  - "Lần 1 và lần 2 được ngửa": $\{HHH, HHT\}$.

Biến cố - Ví dụ (tt)

Tung một đồng xu cho đến khi ra mặt sấp thì dừng $\Omega = \{T, HT, HHT, HHHT, ...\}$

- Biến cố "tung không quá 5 lần": $\{T, HT, HHT, HHHT, HHHHT\}$.
- Biến cố "tung ít nhất 5 lần": $\{HHHHT, HHHHHT, HHHHHHT, ...\}$.
- Biến cố "có hai lần sấp": $\varnothing$.
- Biến cố "có một lần sấp": Ω.

Đo nhiệt độ tại một địa điểm $\Omega = \mathbb{R} = (-\infty, \infty)$ (độ C)

- Biến cố "nước có thể đóng băng": $(-\infty, 0]$.

Đo chiều cao của một người $\Omega = [0, \infty)$ (mét)

- Biến cố "chơi bóng rổ tốt": $[1.8, \infty)$.

Biến cố

"Lý thuyết biến cố" được hình thức hóa bằng "lý thuyết tập hợp". Xét thí nghiệm T với không gian mẫu Ω và các biến cố $E, F \subset \Omega$

- Ω: **biến cố chắc chắn** (certain event),
- $\varnothing$: **biến cố không thể** (impossible event),
- $E^c = \Omega \setminus E$: biến cố **đối** (complement) của E, biến cố "E không xảy ra",
- $E \cup F$: biến cố E **hoặc** (or) F, biến cố "E xảy ra hoặc F xảy ra",
- $E \cap F$: biến cố E **và** (and) F, biến cố "E xảy ra và F xảy ra",
- $E \setminus F$: biến cố E **không** (not) F, biến cố "E xảy ra nhưng F không xảy ra",
- $E \subset F$: E **kéo theo** (imply) F, E xảy ra thì F xảy ra,
- $E = F$: E **là** (is) F, E và F cùng xảy ra hoặc cùng không xảy ra,
- $E \cap F = \varnothing$: E, F **rời nhau** (disjoint) hay **xung khắc** (mutually exclusive), E và F không thể đồng thời xảy ra.

Biến cố - Ví dụ

Tung một xúc xắc, $\Omega = \{1, 2, 3, 4, 5, 6\}$

- Biến cố "được mặt chẵn": $A = \{2, 4, 6\}$.
- Biến cố sơ cấp "được mặt 1": $B = \{1\} = 1$.
- Biến cố "được mặt khác 1": $C = B^c = \Omega \setminus \{1\} = \{2, 3, 4, 5, 6\}$.
- Biến cố "được mặt lẻ": $D = \{1, 3, 5\}$.
- Biến cố A kéo theo biến cố C vì $A \subset C$.
- Biến cố "không được mặt chẵn" và "được mặt lẻ" là như nhau vì $A^c = D$.
- Biến cố "được mặt lẻ nhưng không là 1": $D \setminus B = \{3, 5\}$.
- Biến cố "được mặt chẵn hoặc 1": $A \cup B = \{1, 2, 4, 6\}$.
- Biến cố "được mặt lẻ và khác 1": $D \cap C = D \cap B^c = \{3, 5\}$.
- Biến cố "được mặt chẵn" và "được mặt 1" là xung khắc vì $A \cap B = \varnothing$.

Nội dung

1. Thí nghiệm ngẫu nhiên
2. Biến cố
3. Xác suất
4. Mô hình xác suất đơn giản
5. Công thức cộng xác suất
6. Kỹ thuật đếm
7. Phương pháp xấp xỉ xác suất bằng mô phỏng
8. Ví dụ minh họa - Bài toán sinh nhật

Xác suất

Xét thí nghiệm T với không gian mẫu Ω, một hàm P gắn mỗi biến cố $E \subset \Omega$ với số thực $P(E)$ được gọi là một **độ đo xác suất** (probability measure) trên Ω nếu P thỏa mãn 3 tiên đề:

1. **TĐ1.** Với mọi biến cố $E \subset \Omega, 0 \leq P(E) \leq 1$.
2. **TĐ2.** Với mọi dãy biến cố $E_1, E_2, ...$ đôi một xung khắc ($E_i \cap E_j = \varnothing, \forall i \neq j$):

$$P\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} P(E_i),$$

   tức là $P(E_1 \cup E_2 \cup ...) = P(E_1) + P(E_2) + ...$
3. **TĐ3.** $P(\Omega) = 1$.

$P(E)$ được gọi là **xác suất** (probability) của E và là số đo khả năng xảy ra của biến cố E khi không biết kết quả của thí nghiệm T.

Các cách diễn giải/hiểu/vận dụng xác suất

1. Xác suất là **tỉ lệ** (proportion)
   - Khi khả năng xảy ra của các kết quả là như nhau thì $P(E)$ là tỉ lệ kết quả thuận lợi cho biến cố E.
   - *Ví dụ:* lớp có 20 nữ và 30 nam, gọi ngẫu nhiên một sinh viên trong lớp, xác suất để "nữ được gọi" là 0.4, chính là tỉ lệ nữ của lớp $(20/(20+30))$.
2. Xác suất là **tần suất** (relative frequency)
   - Trong N lần thực hiện lặp lại thí nghiệm T có m lần biến cố E xảy ra thì tần suất xảy ra E là $f(E) = m/N$. Khi N đủ lớn thì $P(E) \approx f(E)$.
   - *Ví dụ:* tung rất nhiều lần một đồng xu không đồng chất thì thấy khoảng 70% số lần là mặt ngửa, vậy xác suất được mặt ngửa là 0.7.
3. Xác suất là **niềm tin** (belief)
   - $P(E)$ là mức độ tin tưởng (của ai đó) việc E sẽ xảy ra khi không biết kết quả của T.
   - *Ví dụ:* theo tôi, xác suất để Việt Nam vô địch World Cup là gần như 0.

Xác suất - Ví dụ

Tung một xúc xắc đồng chất, $\Omega = \{1, 2, 3, 4, 5, 6\}$

$$\begin{cases} P(1) = P(2) = ... = P(6) \text{ (đồng chất)} \\ P(\Omega) = P(1) + P(2) + ... + P(6) \text{ (TĐ2)} \\ P(\Omega) = 1 \text{ (TĐ3)} \end{cases} \Rightarrow P(i) = \frac{1}{6}, i = 1, ..., 6.$$

- Biến cố "được mặt 1": $A = \{1\}$

$$P(A) = P(\{1\}) = \frac{1}{6}.$$

- Biến cố "được mặt chẵn": $B = \{2, 4, 6\}$

$$P(B) = P(\{2\}) + P(\{4\}) + P(\{6\}) = \frac{3}{6} = \frac{1}{2}.$$

Xác suất - Ví dụ (tt)

Tung một xúc xắc không đồng chất với khả năng ra mặt i tỉ lệ với i $(i = 1, ..., 6)$, $\Omega = \{1, 2, 3, 4, 5, 6\}$

$$\begin{cases} P(i) = c \times i, \ c \geq 0 \text{ (TĐ1)} \\ P(\Omega) = \sum_{i=1}^{6} P(i) = c \sum_{i=1}^{6} = 21c \\ P(\Omega) = 1 \end{cases} \quad \Rightarrow 21c = 1 \Rightarrow P(i) = \frac{1}{21} i.$$

- Biến cố "được mặt 1": $A = \{1\}$

$$P(A) = P(\{1\}) = \frac{1}{21}.$$

- Biến cố "được mặt chẵn": $B = \{2, 4, 6\}$

$$P(B) = P(\{2\}) + P(\{4\}) + P(\{6\}) = \frac{2}{21} + \frac{4}{21} + \frac{6}{21} = \frac{12}{21} = \frac{4}{7}.$$

Các tính chất cơ bản của xác suất

1. $P(E^c) = 1 - P(E)$
2. $P(\varnothing) = 0$
3. Nếu $E_1 \subset E_2$ thì $P(E_1) \leq P(E_2)$ và $P(E_2 \setminus E_1) = P(E_2) - P(E_1)$
4. Nếu $E_1 \cap E_2 = \varnothing$ thì $P(E_1 \cup E_2) = P(E_1) + P(E_2)$
5. $P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2)$ (**addition law of probability**)
6. $P(E_1 \cup E_2 \cup E_3) =$
   $P(E_1) + P(E_2) + P(E_3) - P(E_1 \cap E_2) - P(E_1 \cap E_3) - P(E_2 \cap E_3) + P(E_1 \cap E_2 \cap E_3)$
7. $P\left(\bigcup_{i=1}^{\infty} E_i\right) \leq \sum_{i=1}^{\infty} P(E_i)$ (**union bound**)
8. $P\left(\bigcap_{i=1}^{\infty} E_i\right) \geq 1 - \sum_{i=1}^{\infty} P(E_i^c)$ (**Bonferroni inequality**)

Các tính chất cơ bản của xác suất (tt)

Chứng minh rằng $P(E^c) = 1 - P(E)$ với mọi biến cố $E \subset \Omega$.

Chứng minh. Xét $E \subset \Omega$ là một biến cố bất kì, vì

$$E \cup E^c = \Omega \tag{1}$$

và

$$E \cap E^c = \varnothing \tag{2}$$

nên ta có

$$1 \overset{\text{TĐ3}}{=} P(\Omega) \overset{(1)}{=} P(E \cup E^c) \overset{(2) \text{ và } \text{TĐ2}}{=} P(E) + P(E^c).$$

Suy ra

$$P(E^c) = 1 - P(E).$$

Nội dung

1. Thí nghiệm ngẫu nhiên
2. Biến cố
3. Xác suất
4. Mô hình xác suất đơn giản
5. Công thức cộng xác suất
6. Kỹ thuật đếm
7. Phương pháp xấp xỉ xác suất bằng mô phỏng
8. Ví dụ minh họa - Bài toán sinh nhật

Mô hình xác suất đơn giản

Khi không gian mẫu hữu hạn, $\Omega = \{\omega_1, \omega_2, ..., \omega_n\}$, độ đo xác suất được xác định bởi xác suất của các biến cố sơ cấp $p_i = P(\omega_i)$

- $p_i \geq 0, i = 1, ..., n,$
- $\sum_{i=1}^{n} p_i = 1,$
- $P(E) = \sum_{w_i \in E} p_i$ với mọi biến cố $E \subset \Omega$.

Khi không gian mẫu hữu hạn và các **kết quả đồng khả năng** (equiprobable outcomes), ta có **mô hình xác suất đơn giản** (simple/classical probability model)

- $p_i = \frac{1}{n}, i = 1, ..., n,$
- $P(E) = \frac{|E|}{|\Omega|}$ với mọi biến cố $E \subset \Omega$.
- Xác suất là tỉ lệ và việc tính xác suất được đưa về việc **đếm** (counting).

Mô hình xác suất đơn giản - Ví dụ

Tung một đồng xu đồng chất 3 lần

- $\Omega = \{H, T\}^3, |\Omega| = 8,$
- Biến cố "được đúng 2 lần ngửa": $E = \{HHT, HTH, THH\}$

$$P(E) = \frac{|E|}{|\Omega|} = \frac{3}{8}.$$

Tung một đồng xu đồng chất 3 lần

- $\Omega = \{0, 1, 2, 3\}$ (số lần được ngửa), $|\Omega| = 4,$
- Biến cố "được đúng 2 lần ngửa": $E = \{2\}$

$$P(E) = \frac{|E|}{|\Omega|} = \frac{1}{4}.$$

- Lí luận sai: vì các biến cố sơ cấp không đồng khả năng nên không phải là mô hình xác suất đơn giản.

Nội dung

1. Thí nghiệm ngẫu nhiên
2. Biến cố
3. Xác suất
4. Mô hình xác suất đơn giản
5. **Công thức cộng xác suất**
6. Kỹ thuật đếm
7. Phương pháp xấp xỉ xác suất bằng mô phỏng
8. Ví dụ minh họa - Bài toán sinh nhật

Công thức cộng xác suất

1. Nếu $A_1, A_2, ..., A_n$ rời nhau $(A_i \cap A_j = \varnothing, \forall i \neq j)$ thì

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i)$$

2. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
3. $P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$
4. $P(\cup_{i=1}^{n} A_i) = \sum_i P(A_i) - \sum_{i<j} P(A_i \cap A_j) + \sum_{i<j<k} P(A_i \cap A_j \cap A_k) - \ \ldots \ + (-1)^{n+1} P(A_1 \cap A_2 \cap ... \cap A_n)$

Công thức cộng xác suất - Ví dụ

Một công ty sản xuất 130 bóng đèn với "phân bố" chất lượng dựa trên 2 tiêu chí là kiểu dáng và độ sáng như bảng sau

		Kiểu dáng		
		Đạt	Không đạt	**Tổng**
Độ sáng	Đạt	117	3	120
	Không đạt	8	2	10
	Tổng	125	5	130

Chọn ngẫu nhiên 1 bóng đèn. Tính xác suất bóng đèn được chọn thỏa ít nhất một trong hai tiêu chí trên?

Công thức cộng xác suất - Ví dụ (tt)

Mô hình xác suất đơn giản, $|\Omega| = 130$.
Đặt các biến cố

- A: "bóng đèn được chọn thỏa tiêu chí kiểu dáng",
- B: "bóng đèn được chọn thỏa tiêu chí độ sáng".

Xác suất bóng đèn được chọn thỏa ít nhất một trong hai tiêu chí trên là

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = \frac{125}{130} + \frac{120}{130} - \frac{117}{130} = \frac{128}{130}.$$

Cách khác (dùng "công thức trừ")

$$P(A \cup B) = 1 - P\left((A \cup B)^c\right) = 1 - P(A^c \cap B^c) = 1 - \frac{2}{130} = \frac{128}{130}.$$

Nội dung

1. Thí nghiệm ngẫu nhiên
2. Biến cố
3. Xác suất
4. Mô hình xác suất đơn giản
5. Công thức cộng xác suất
6. Kỹ thuật đếm
7. Phương pháp xấp xỉ xác suất bằng mô phỏng
8. Ví dụ minh họa - Bài toán sinh nhật

Qui tắc nhân (multiplication rule)

Nếu một việc T được thực hiện bằng 2 bước độc lập A, B; có m cách thực hiện A và n cách thực hiện B thì có $m \times n$ cách thực hiện T

- $T = A \times B = \{(x, y) : x \in A, y \in B\} \Rightarrow |T| = |A| \times |B|$.
- $|A_1 \times A_2 \times ... \times A_k| = |A_1| \times |A_2| \times ... \times |A_k|$.
- $|A^k| = |A|^k$.

Ví dụ: tung một xúc xắc 3 lần

$$\Omega = \{(i, j, k) : i, j, k \in \{1, 2, 3, 4, 5, 6\}\} = \{1, ..., 6\}^3$$

- $|\Omega| = 6^3 = 216$,
- Biến cố "được tổng cộng 4 nút": $A = \{(1, 1, 2), (1, 2, 1), (2, 1, 1)\}$

$$P(A) = \frac{|A|}{|\Omega|} = \frac{3}{216} = \frac{1}{72}.$$

Qui tắc nhân (tt)

Lấy mẫu có hoàn lại (sampling with replacement)

- Từ tập A có n phần tử, chọn lần lượt k lần, mỗi lần một phần tử có hoàn lại.
- Số kết quả chọn, có kể đến thứ tự, là n^k.

Ví dụ: một hộp có 3 bi xanh và 2 bi đỏ, bốc ngẫu nhiên 3 lần có hoàn lại

- $|\Omega| = 5^3$,
- Biến cố A: "cả 3 lần bốc đều được bi đỏ" có xác suất

$$P(A) = \frac{|A|}{|\Omega|} = \frac{2^3}{5^3} = 0.064.$$

Hoán vị (permutation)

Lấy mẫu không hoàn lại (sampling without replacement)

- Từ tập A có n phần tử, chọn ra lần lượt k phần tử không hoàn lại. Mỗi kết quả chọn, có kể đến thứ tự, được gọi là một **chỉnh hợp** chọn k của n phần tử. Số chỉnh hợp chọn k của n phần tử là

$$P_n^k = n \times (n-1) \times ... \times (n-k+1) = \frac{n!}{(n-k)!}.$$

- Một chỉnh hợp chọn n của n phần tử được gọi là một **hoán vị** của n phần tử. Số hoán vị của n phần tử là $P_n = n!$.

Ví dụ: chọn ngẫu nhiên 3 người từ 10 người (trong đó có Bình) để trao giải nhất, nhì, ba, $|\Omega| = P_{10}^3$, biến cố A: "Bình được giải nhất" có xác suất

$$P(A) = \frac{|A|}{|\Omega|} = \frac{P_9^2}{P_{10}^3} = 0.1.$$

Tổ hợp (combination)

Tổ hợp (combination)

- Từ tập A có n phần tử, chọn ra k phần tử không hoàn lại. Mỗi kết quả chọn, không kể thứ tự, được gọi là một **tổ hợp** chọn k của n phần tử.
- Số tổ hợp chọn chọn k của n phần tử là

$$C_n^k = \frac{n!}{k!(n-k)!}.$$

Ví dụ: tung một đồng xu 10 lần, $\Omega = \{H, T\}^{10}, |\Omega| = 2^{10}$

- Biến cố A: "được 3 lần ngửa",
- Mỗi kết quả thuận lợi cho A sẽ có mặt ngửa trong 3 lần chọn từ 10 lần, như vậy $|A| = C_{10}^3$ và

$$P(A) = \frac{|A|}{|\Omega|} = \frac{C_{10}^3}{2^{10}} = 0.1172.$$

Kỹ thuật đếm - Ví dụ

Một lớp có 15 nam và 30 nữ. Chọn ra ngẫu nhiên 10 học sinh đi lao động. Tính xác suất có 3 nam được chọn.

- Kết quả của thí nghiệm là một tổ hợp chọn 10 của 45 học sinh: $|\Omega| = C_{45}^{10}$
- Do chọn ngẫu nhiên nên ta có mô hình xác suất đơn giản
- Gọi A là biến cố có 3 nam được chọn. Kết quả thuận lợi cho A là lựa chọn gồm 3 nam và 7 nữ
- Có C_{15}^{3} cách chọn 3 nam từ 15 nam và C_{30}^{7} cách chọn 7 nữ từ 30 nữ
- Theo qui tắc nhân ta có số kết quả thuận lợi cho A là: $|A| = C_{15}^{3} \times C_{30}^{7}$
- Xác suất có 3 nam được chọn là

$$P(A) = \frac{|A|}{|\Omega|} = \frac{C_{15}^{3} \times C_{30}^{7}}{C_{45}^{10}} = 0.2904.$$

Nội dung

1. Thí nghiệm ngẫu nhiên
2. Biến cố
3. Xác suất
4. Mô hình xác suất đơn giản
5. Công thức cộng xác suất
6. Kỹ thuật đếm
7. **Phương pháp xấp xỉ xác suất bằng mô phỏng**
8. Ví dụ minh họa - Bài toán sinh nhật

Phương pháp xấp xỉ xác suất bằng mô phỏng

Để xấp xỉ xác suất của một biến cố E liên quan đến thí nghiệm T, ta có thể dùng phương pháp thống kê như sau

- Thực hiện lặp lại N lần (độc lập) thí nghiệm T và đếm số lần biến cố E xảy ra, m. Khi đó $f(E) = \frac{m}{N}$ được gọi là tần suất của E.
- Khi N đủ lớn, ta có $f(E) \approx P(E)$.
- Việc thực hiện lặp lại nhiều lần thí nghiệm T có thể được **mô phỏng** (simulation) trên máy tính.

Nội dung

1. Thí nghiệm ngẫu nhiên
2. Biến cố
3. Xác suất
4. Mô hình xác suất đơn giản
5. Công thức cộng xác suất
6. Kỹ thuật đếm
7. Phương pháp xấp xỉ xác suất bằng mô phỏng
8. **Ví dụ minh họa - Bài toán sinh nhật**

Bài toán sinh nhật - Tính toán chính xác

Bài toán sinh nhật (the birthday problem). Tính xác suất p của biến cố có ít nhất 2 người cùng sinh nhật (cùng ngày và tháng sinh) trong nhóm k người?
Giả sử: ngày sinh của mỗi người là một ngày ngẫu nhiên trong một năm gồm 365 ngày và "không liên quan nhau".

Giải. Không gian mẫu $\Omega = \{1, 2, ..., 365\}^k$ với $|\Omega| = 365^k$.
Đặt các biến cố:

- A là "có ít nhất 2 người cùng sinh nhật",
- B là "không có người nào cùng sinh nhật".

Như vậy, $A = B^c$ và $B = \{\text{chỉnh hợp chọn } k \text{ của } \{1, 2, ..., 365\}\}$ với $|B| = P_{365}^k$.
Dùng mô hình xác suất đơn giản ta có

$$p = P(A) = 1 - P(B) = 1 - \frac{P_{365}^k}{365^k} = 1 - \frac{365!}{(365-k)!365^k}$$

Bài toán sinh nhật - Mô phỏng

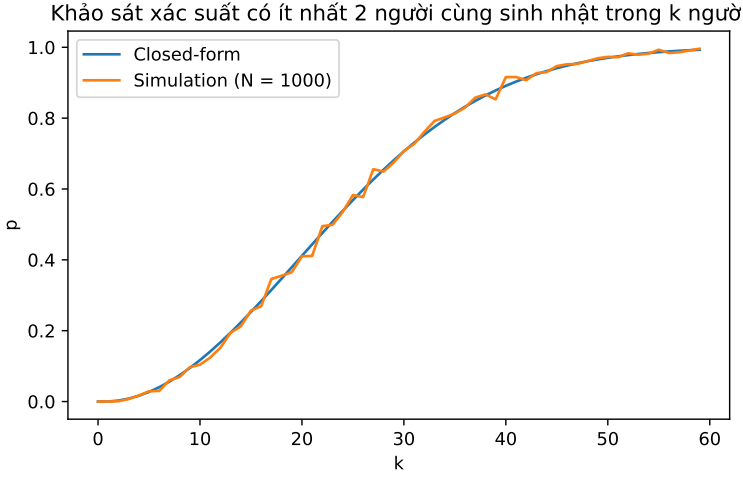
```
def birthday(k):
    return [random.randint(1, 365) for _ in range (k)]

def at_least_2(outcome):
    return len(set(outcome)) < len(outcome)

def relative_frequency(k, N, event):
    m = sum(event(birthday(k)) for _ in range (N))
    return m/N

relative_frequency(50 , 1000, at_least_2)
#0.973
```

Bài toán sinh nhật - Kết quả

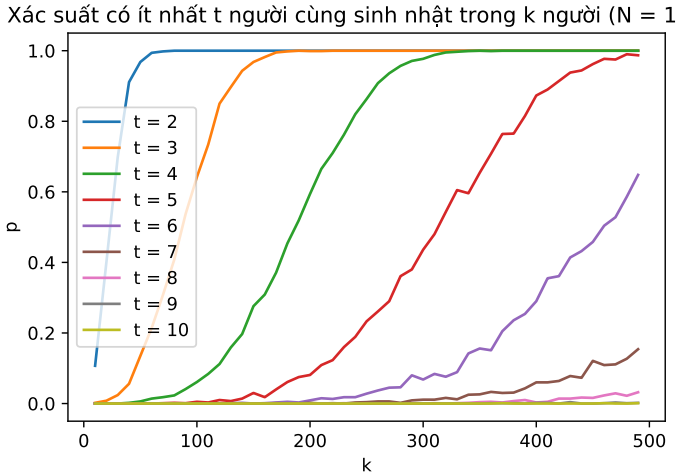


Bài toán sinh nhật - Mở rộng

Mở rộng bài toán. Tính xác suất p của biến cố có ít nhất t người cùng sinh nhật (cùng ngày và tháng sinh) trong nhóm k người?

- Tính toán lý thuyết: được kết quả chính xác nhưng khó và không khả thi trong nhiều trường hợp.
- Mô phỏng máy tính: dễ dàng (chỉnh sửa mã) nhưng tốn thời gian (đợi mô phỏng), tài nguyên máy và chỉ được kết quả xấp xỉ.
  (Xem mã Python trong Notebook)

Bài toán sinh nhật - Mở rộng (tt)



Tài liệu tham khảo

Chapter 1. Morris H. DeGroot, Mark J. Schervish. *Probability and Statistics*. Addison-Wesley, 2012.

Chapter 1, 2. H. Pishro-Nik. *"Introduction to probability, statistics, and random processes"*, available at `https://www.probabilitycourse.com`. Kappa Research LLC, 2014.